

Abstract

Apparatus for building a stochastic model of a time sequential data sequence, the data sequence comprising symbols selected from a finite symbol set, the apparatus comprising: an input for receiving said data sequence, a tree builder for expressing said symbols as a series of counters within nodes, each node having a counter for each symbol, each node having a position within said tree, said position expressing a symbol sequence and each counter indicating a number of its corresponding symbol which follows a symbol sequence of its respective node, and a tree reducer for reducing said tree to an irreducible set of conditional probabilities of relationships between symbols in said input data sequence. The tree may then be used to carry out a comparison with a new data sequence to determine a statistical distance between the old and the new data sequence.